

## New developments

### **Herbert Baumann, Reiner Dirksmeyer, Peter Wittenburg - Long-Term Archiving**

It was reported frequently that two aspects are important to increase the chance of a survival of the bit-stream representations of the material we are storing about languages and music traditions that will be extinct soon. (1) The data has to be migrated frequently to guarantee that state-of-the-art storage media are used that are fully supported by hardware and software. (2) The data has to be copied and distributed to cope with all kinds of risks – even political ones - that could destroy the storage media used. The MPI team has finished its activity to have at least 5 copies of the DOBES data. Two copies are automatically created in the MPI storage system (RAID Disk Array and Tape Library). A third copy is stored on a standard PC system having a large RAID Disk Array that is within the control of the MPI team, but located in another building.

A fourth copy is transferred to the computer center of the Max-Planck-Society in Göttingen (GWDG) by using the RSYNC protocol provided for example in standard UNIX systems. The transfer is initiated by the GWDG, the protocol is efficient but lacks modern encryption capabilities. To achieve the full transmission speed of 5 MByte/sec five sessions are started in parallel. A fifth copy was generated in the mean time at the other computer center of the Max-Planck-Society in Munich (RZG). Here the well-known Andrew File System (AFS) is used as protocol. At the MPI an AFS client was installed that establishes connections with the AFS server in Munich, i.e., the transfer is initiated by the archivist. AFS makes use of state-of-the-art authentication and encryption. Also here several channels are opened in parallel to achieve the full 2.5 MByte/sec exchange speed.

Both procedures guarantee that at regular intervals the changes in our DOBES archive are synchronized with the two computer centers. At both centers, GWDG and RZG, local strategies are applied to maintain several copies of all stored data in different buildings, i.e., all DOBES data is now stored in at least 7 different storage systems. The DOBES archivist sees it as an advantage that two different protocols are applied and that the two centers use different storage technologies. Currently, the Max-Planck-Society discusses at a high level what kind of guarantees can be given to the institutions for long-term

storage support for their data, since many disciplines share this fundamental problem.

In the committees dealing with this question of long-term preservation it is consensus that guaranteeing the interpretation of the bit-streams is a task of the community and not the centers. Adherence to open standards and organizational, encoding and format coherence will be relevant criteria to determine the chance that the data will be migrated in time to state-of-the-art representation standards.